Patent claims

1.   A method for computer-aided determination of a
sequence of actions for a system which has states, a
transition in state between two states being performed
on the basis of an action, in the case of which the
determination of the sequence of actions is performed
in such a way that a sequence of states resulting from
the sequence of actions is optimized with regard to a
prescribed optimization function, the optimization
function including a variable parameter with the aid of
which it is possible to set a risk which the resulting
sequence of states has with respect to a prescribed
state of the system.

2.   The method as claimed in claim 1, in which a
method of approximative dynamic programming is used for
the purpose of determination.

3.   The method as claimed in claim 2, in which the
method of approximative dynamic programming is a method
based on Q-learning.

4.   The method as claimed in claim 3, in which the
optimization function OFQ is formed within Q-learning
in accordance with the following rule:

$$OFQ = Q\left(x; w^a\right),$$

- ● x denoting a state in a state space X
- ● a denoting an action from an action space A, and
- ● w^a denoting the weights of a function approximator
  which belong to the action a,

and in which the weights of the function approximator
are adapted in accordance with the following rule:

$$w^{a_t}_{t+1} = w^{a_t}_t + \eta_t \cdot \aleph^\kappa(d_t) \cdot \nabla Q\left(x_t; w^{a_t}_t\right)$$

with the abbreviation

$$d_t = r\left(x_t, a_t, x_{t+1}\right) + \gamma \max_{a \in A} Q\left(x_{t+1}, w^a_t\right) - Q\left(x_t, w^{a_t}_t\right)$$

- x_t, x_t+1 respectively denoting a state in the state space X,
- a_t denoting an action from an action space A,
- γ denoting a prescribable reduction factor,
- $w^{a_t}_t$ denoting the weighting vector associated with the action a_t before the adaptation step,
- $w^{a_t}_{t+1}$ denoting the weighing vector associated with the action a_t after the adaptation step,
- η_t (t = 1, ...) denoting a prescribable step size sequence,
- $\kappa \in [-1; 1]$ denoting a risk monitoring parameter,
- $\aleph^\kappa$ denoting a risk monitoring function $\aleph^\kappa(\xi) = (1 - \kappa\,\mathrm{sign}(\xi))\xi$,
- $\nabla Q(\cdot; \cdot)$ denoting the derivation of the function approximator according to its weights, and
- $r(x_t, a_t, x_{t+1})$ denoting a gain upon the transition of state from the state x_t to the subsequent state x_{t+1}.

5. The method as claimed in claim 2, in which the method of approximative dynamic programming is a method based on TD(λ)-learning.

6. The method as claimed in claim 5, in which the optimization function OFTD is formed within TD(λ)-learning in accordance with the following rule:

OFTD = $J(x;w)$

- •⠀⠀x denoting a state in a state space X,
- •⠀⠀a denoting an action from an action space A, and
- •⠀⠀w denoting the weights of a function approximator

and in which the weights of the function approximator are adapted in accordance with the following rule:

$$w_{t+1} = w_t + \eta_t \cdot \aleph^\kappa(d_t) \cdot z_t$$

with the abbreviations

$$d_t = r(w_t, a_t, x_{t+1}) + \gamma J(x_{t+1}; w_t) - J(x_t; w_t),$$

$$z_t = \lambda \cdot \gamma \cdot z_{t-1} + \nabla J(x_t; w_t),$$

$$z_{-1} = 0$$

- •⠀⠀x_t, x_{t+1} respectively denoting a state in the state space X,
- •⠀⠀a_t denoting an action from an action space A,
- •⠀⠀γ denoting a prescribable reduction factor,
- •⠀⠀w_t denoting the weighting vector before the adaptation step,
- •⠀⠀w_{t+1} denoting the weighting vector after the adaptation step,
- •⠀⠀η_t (t = 1, ...) denoting a prescribable step size sequence,
- •⠀⠀$\kappa \in [-1; 1]$ denoting a risk monitoring parameter,
- •⠀⠀$\aleph^\kappa$ denoting a risk monitoring function $\aleph^\kappa (\xi) = (1 - \kappa \operatorname{sign}(\xi))\xi$,
- •⠀⠀$\nabla J(\cdot; \cdot)$ denoting the derivation of the function approximator according to its weights, and
- •⠀⠀$r(x_t, a_t, x_{t+1})$ denoting a gain upon the transition of state from the state x_t to the subsequent state x_{t+1}

7.   The method as claimed in one of claims 1 to 6, in which the system is a technical system of which before the determination measured values are measured which are used in determining the sequence of actions.

5   8.   The method as claimed in claim 7, in which the technical system is subjected to open-loop control in accordance with the sequence of actions.

9.   The method as claimed in claim 7, in which the technical system is subjected to closed-loop control in 10 accordance with the sequence of actions.

10.   The method as claimed in one of claims 1 to 9, in which the system is modeled as a Markov decision problem.

11.   The method as claimed in one of claims 1 to 10, 15 being used in a traffic management system.

12.   The method as claimed in one of claims 1 to 10, being used in a communications system.

13.   The method as claimed in one of claims 1 to 10, being used to carry out access control in a 20 communications network.

14.   The method as claimed in one of claims 1 to 10, being used to carry out a routing in a communications network.

15.   An arrangement for determining a sequence of 25 actions for a system which has states, a transition in state between two states being performed on the basis of an action,

having a processor which is set up in such a way that the determination of the sequence of actions can be performed in such a way that a sequence of states resulting from the sequence of actions is optimized

5   with regard to a prescribed optimization function, the optimization function including a variable parameter with the aid of which it is possible to set a risk which the resulting sequence of states has with respect to a prescribed state of the system.

10   16.   The arrangement as claimed in claim 15, being used to subject a technical system to open-loop control.

17.   The arrangement as claimed in claim 15, being used to subject a technical system to closed-loop

15   control.

18.   The arrangement as claimed in claim 15, being used in a traffic management system.

19.   The arrangement as claimed in claim 15, being used in a communications system.

20   20.   The arrangement as claimed in claim 15, being used to carry out access control in a communications network.

21.   The arrangement as claimed in claim 15, being used to carry out a routing in a communications

25   network.